AD 488273

Additional Notes On The Negative Binomial Distribution

by

Dalton H. Wright

ALRAND Rpt 26
SPCC Mechanicsburg, Pa.
9 February 1961

THE NEGATIVE BINOMIAL DISTRIBUTION

Part 1

There are only two parameters to be estimated from the sample data when the generalized Negative Binomial frequency function is used to represent the population: (1) $E(x)$ and (2) σ_x^2. The formula for this frequency function according to Hildebrand* is:

$$f(x) = \frac{k(k+1)\ldots\ldots(k+x-1)}{1\cdot 2\cdot 3\ldots\ldots x}\left(\frac{1-q}{q}\right)^x\left(\frac{1}{q}\right)^k, \quad x = 1, 2\ldots$$

$$f(0) = \left(\frac{1}{q}\right)^k \quad \text{where } p = q - 1 \text{ and}$$

$$q = \frac{\sigma_x^2}{E(x)}, \quad k = \frac{E(x)}{q-1}, \quad q > 1, \; k > 0, \; x \geq 0, \; E(x) = kp, \text{ and}$$

$$\sigma_x^2 = kpq.$$

A useful computing formula is given by the relationship $f(x) = \frac{k+x-1}{x}\left(\frac{p}{q}\right) f(x-1)$. The values of q and k are computed from $E(x)$ and σ_x^2. Knowing q and k the theoretical probabilities are generated by allowing x to take values $0, 1, 2, \ldots \infty$. The probabilities should then be compared with the observed frequencies or histogram by using the χ^2 test or the k - s test.

*See ALRAND Rpt 19

Part II

We now ask ourselves why does the Negative Binomial frequency function characterize so many demand analysis problems and when is it safe to <u>assume</u> the demand variable is distributed according to this frequency function. In cases where very little data is available an assumption regarding the type of frequency function is often necessary. One reason this frequency function is so useful is its great flexibility due to the two parameters in the formula. It is not a single formula but a two family set of curves. However, our main interest is the derivation of the frequency function. This should suggest why and <u>when</u> the frequency function is a proper mathematical model for demand analysis. The ordinary Poisson frequency function is expected to be useful in demand analysis because in its derivation the main assumption is that the values of the variable are randomly spaced over time. Then the frequency function of the variable $x(t)$ or demand will have an ordinary Poisson frequency function. This assumption is apparently not valid for a large variance in demand due to clustering where the clusters are not random and hence the demand will not have a Poisson frequency function in this case.

The Negative Binomial or Pascal frequency function is defined by

$f(k; r, p) = \binom{r+k-1}{k} \cdot p^r q^k$, $k = 0, 1, 2, \ldots.$ It equals the probability that $r + k$ Bernoulli trials are needed to produce exactly r successes. The probability of a unit demand occurring in a given time interval is denoted by p where $p + q = 1$. The frequency function, $f(k; r, p)$, may be interpreted as the probability distribution for the waiting time to the r^{th} "success." This frequency function is related to the Poisson frequency function in the following manner:

$$f(k; r, p) \rightarrow \frac{e^{-\lambda} \lambda^k}{k!} \quad \text{if } q \rightarrow 0, \; r \rightarrow \infty, \text{ and } rq = \lambda \text{ remains}$$

fixed.

The frequency function, $f(k, r, p)$, is called Negative Binomial due to the following relationship:

$$\sum_{k=0}^{\infty} f(k; r, p) = \sum_{k=0}^{\infty} \binom{r+k-1}{k} p^r q^k = \sum_{k=0}^{\infty} \binom{-r}{k} p^r (-q)^k$$

$$= p^r \sum_{k=0}^{\infty} \binom{-r}{k} (-q)^k = p^r (1-q)^{-r} = 1 \quad \text{where the expansion}$$

of $p^r(1-q)^{-r}$ by the binomial theorem generates the probabilities $f(k; r, p)$.

Part III

Let n be a random variable having a Poisson frequency function with probabilities $\frac{\mu^n e^{-\mu}}{n!}$, where n = 0, 1, ... ∞. If we let μ be a random variable with the frequency function $\frac{\alpha^\lambda}{\Gamma(\lambda)} \mu^{\lambda - 1} e^{-\alpha\mu}$ the probability that n takes any given ν is given by:

$$\int_0^\infty \frac{\mu^\nu}{\nu!} e^{-\mu} \cdot \frac{\alpha^\lambda}{\Gamma(\lambda)} \mu^{\lambda - 1} e^{-\alpha\mu} d\mu =$$

$$\left(\frac{\alpha}{1 + \alpha}\right)^\lambda \binom{-\lambda}{\nu} \frac{(-1)^\nu}{(1 + \alpha)^\nu}$$

This frequency function is the Negative Binomial frequency function. This derivation of the Negative Binomial might suggest some explanation why the Negative Binomial is more useful than the ordinary Poisson for certain values of the variance in demand analysis. It has interesting applications to accident and sickness statistics.

Part IV

In the Negative Binomial distribution,

$$\begin{cases} f(x, \mu, \sigma_x^2) = \dfrac{k(k+1)\ldots(k+x-1)}{x!} \cdot \left(\dfrac{1-q}{q}\right)^x \dfrac{1}{q^k}, \quad x = 1, 2, \ldots . \\ f(0) = \left(\dfrac{1}{q}\right)^k \quad \text{where } p = q - 1, \quad q = \sigma_x^2/E(x), \end{cases}$$

$k = \dfrac{E(x)}{q-1}$, $q > 1$, $k > 0$, $x \geq 0$, $E(x) = kp$ and

$\sigma_x^2 = kpq$ and $M_x(\theta) = M_x(\theta) = (q - pe^{\theta})^{-k}$, let

$\sigma_x^2 \to E(x)$ or $q \to 1$.

Then the Negative Binomial distribution approaches the Poisson distribution as $q \to 1$.

Proof: The moment generating function for the Negative Binomial distribution is $M_x(\theta) = (q - pe^{\theta})^{-k}$. Taking the logarithm of both members of this equation and performing some calculations we have:

$$\begin{aligned} \log M_x(\theta) &= -k \log (q - pe^{\theta}) \\ &= -k \log \left[1 + (p - pe^{\theta})\right] \\ &= -k \left\{(p - pe^{\theta}) - 1/2(p - pe^{\theta})^2 + 1/3(p - pe^{\theta})^3 - \ldots .\right\} \\ &= -\frac{\mu}{q-1} \left\{p(1 - e^{\theta}) - 1/2p^2(1 - e^{\theta})^2 + 1/3p^3(1 - e^{\theta})^3 - \ldots .\right\} \\ &= -\mu\left\{(1 - e^{\theta}) - 1/2(q-1)(1 - e^{\theta})^2 + 1/3(q-1)^2 \right. \\ &\qquad \left. \cdot (1 - e^{\theta})^3 - \ldots .\right\} \end{aligned}$$

Letting $q \to 1$ and μ remain fixed we have: $\log M_x(\theta) \xrightarrow[q \to 1]{} -\mu(1 - e^{\theta})$ or in exponential form $M_x(\theta) \xrightarrow[q \to 1]{} e^{-\mu(1 - e^{\theta})}$. But this is the moment generating function for the ordinary Poisson distribution. Therefore, the Negative Binomial distribution approaches the Poisson distribution as $q \to 1$. This theorem has important implications in demand analysis at SPCC. The Poisson distribution is a special case of the Negative Binomial distribution--namely when $q \to 1$. When q is slightly greater than or equal to one, we do not gain anything by switching to the Poisson distribution unless the computing becomes simpler. This theorem also lends support to the suggestion that the demand variable (usage rate) is usually an ordinary Poisson variable which has been distorted due to present requisitioning procedures.

By reference to present requisitioning procedures I have in mind suggestions made by CDR Mills and others regarding the effects of order sizes and ordering habits. Large order sizes greatly increase the variance and have no effect on the mean of the demands over a long time period. Another interesting requisitioning habit is that apparently people prefer to order in even numbered units rather than odd numbered units. This strange behavior, if true, increases the variance of the demands with no effect on the mean.

There are two approaches to these problems: (1) accept the demand rates as they are and analyze them and (2) try to improve the requisitioning procedures.

The following table illustrates the relationship between the Poisson distribution and the Negative Binomial distribution.

Poisson $\mu = .5$		Negative Binomial $\mu = .5$ $q = 1.01$ $k = 50.0$	
x	f(x)	x	f(x)
0	.60653	0	.60835
1	.30327	1	.30113
2	.07582	2	.07602
3	.01264	3	.01305
4	.00158	4	.00171